MEDLARS: Performance, Problems, Possibilities

By Scott Adams, *Deputy Director*

National Library of Medicine
Bethesda, Maryland

ABSTRACT

Operational experience during the first year of MEDLARS is described in relation to the Library's primary objectives of index publication, recurring bibliographies, and demand search services. Progress toward the implementation of two secondary objectives—the input of cataloging copy and the decentralization of the system—is also reported. MEDLARS has been intensively tested during its first operational year, its strengths and weaknesses have been identified, and it is now prepared to develop its service potential.

In reporting to the medical library profession at the Second International Congress on Medical Librarianship, the National Library of Medicine predicted some of the problems which it subsequently experienced in inaugurating MEDLARS:

> The scheduling of successive issues [of the *Index Medicus*] and the elimination of errors will be priority responsibilities, and staff time available for experimental production of other services will be contingent on the operational success of the computerized *Index Medicus*.... we shall not be in a position to offer machine search services widely because of our preoccupation with learning to understand the system, and making it operate efficiently (1).

The first year of MEDLARS operation has proved to be one of intensive trial, test, experiment, evaluation, and change. Internal and external pressures alike have been brought to bear on the system, forcing the Library to extend the capabilities of a limited staff. MEDLARS has been highly conspicuous nationally and internationally, and the variety of challenge and the Library's necessarily experimental response have made for an extremely busy year.

The following may be considered as a summary report of some of the more significant developments in the use of MEDLARS since January 1964.

Chronology

The contract with the General Electric Company called for the delivery to the Library of an operational system on January 31, 1964. Operation

by the Library staff began in February 1964. However, a contract relationship with GE was continued in order to complete work on specialized projects such as the GRACE photoprinter. All during the calendar year 1963 the Library operated a two-track system, preparing copy for the 1963 issues of the *Index Medicus* through the Listomatic camera and concurrently converting the same material for MEDLARS input. Personnel resources were strained, and backlogs developed. Consequently the Library made a decision to eliminate from MEDLARS input all foreign-language citations of 1963 and concentrated on those in the English language. As a result, approximately 45,000 citations representing English-language articles published in 1963 were available on magnetic tape by January 1, 1964. Inclusion of foreign-language citations was initiated in January of 1964.

During the fall of 1963 the Library ran test demand searches in order to test the limitations of the MEDLARS system. Approximately 300 test searches had been accomplished successfully by January 1, 1964. Also during the fall of 1963, an increasing number of engineering problems relating to the development of GRACE forced the Library to a decision to produce the first issues of the *Index Medicus* for 1964 on a machine printer as an alternate composing device. As of January 1, 1964, therefore, the Library had the assurance of producing the *Index Medicus* through MEDLARS, albeit with less than satisfactory typography, had some experience in demand search procedures, and was in the early stages of discussion of the forms and formats of recurring bibliographies.

Index Medicus

The January issue of *Index Medicus* contained 5,121 citations in 381 pages. It was run using a Honeywell chain-drive printer at the Army Map Service, operating at a speed of 600 lines per minute. By April the *Index Medicus* included 13,040 citations in 977 pages, producing an issue two inches thick. While users were struggling with this increasing volume of solid capital letter printout, the GE Company and Photon, Inc., intensified efforts to perfect GRACE. As a safeguard against the possibility of further delays, the Library, with the assistance of the IBM Corporation and Honeywell, Inc., engaged in a crash programming effort to make it possible to compose the *Index Medicus* through the IBM 1403 computer-printer, using upper and lower case characters. GRACE was delivered to the Library on May 11, 1964, and a month of testing followed. The Library decided to use the 1403 program for the July issue, partly to acknowledge the cooperation rendered by IBM and Honeywell in achieving an improved printing quality and partly to demonstrate and gain experience with the alternate composing device.

The first issue of the *Index Medicus* to be composed by GRACE, therefore, was that of August 1964. The run was accomplished in sixteen hours. The system detected on the average only one missing letter in ninety pages of composition, and the operational success of GRACE fully demonstrated the feasibility of high-speed photocomposition for the *Index Medicus*. The significance of GRACE as a phenomenon in the printing industry is just beginning to be appreciated. It is the fastest phototypesetter in the world, and the Library has held several demonstrations for representatives of the printing and book publishing industry.

With the August issue, the production of the *Index Medicus* through the MEDLARS system became stabilized. No major breakdowns have occurred. The "List of Journals Indexed" and the 1965 edition of *MeSH* accompanying the January 1965 issue were also prepared through GRACE, and it has been decided to include as a special section in the *Index Medicus* the monthly selection of critical review papers which will be cumulated annually to form the *Bibliography of Medical Reviews*.

Abridged Index Medicus

The desirability of an *Abridged Index Medicus* had been discussed at regional meetings of the Medical Library Association with inconclusive results. In the interest of testing the need of practicing physicians for a selective index, the Library borrowed a page from its past and prepared a small edition of a Specimen Fasciculus of an *Abridged Index Medicus* in December 1964. This sample contains 3,667 citations from a month's input of 219 journal titles, 98.2% of which are in the English language and clinically oriented. Produced through demand search procedures and composed by GRACE, this experimental issue will be used to test the market for a selective index.

Demand Searches

As mentioned above, the Library ran a number of demand search requests in the fall of 1963 under what might be called laboratory conditions. The Library stated in the January 1964 issue of the *NLM News* its intention to make search services available. Concurrently it solicited trial of the demand search service by research scientists located at the National Institutes of Health and other local institutions. The results of these tests indicated two major difficulties. First, the time necessary to process search requests and the limited number of staff available made it impossible for the Library to accommodate any large volume of search requests. Second, weaknesses in *MeSH* were responsible for the retrieval

of a number of papers considered nonrelevant to the purposes of the research scientists.

Accordingly, the Library adopted a policy of providing search services only under test conditions, while it initiated a major effort to strengthen its indexing performance and *MeSH* so that they might more accurately reflect and be responsive to the needs of research scientists. By test conditions, the Library stipulates a willingness to review and evaluate the search results, noting those considered relevant to the question and suggesting reasons for nonrelevance of others. This critical feedback has constituted a *sine qua non* of demand search service during the course of the year. Although searches have been performed experimentally for many nonlocal scientists, they have been for the most part individually negotiated and performed for the purpose of evaluation and systems improvement, rather than service.

To date, the Library has accomplished 577 searches in addition to the 300 related to systems acceptance. A few examples of recent searches of the 1963 English-language and 1964 literatures follow:

Effects of chelating agents on tumor and virus growth	77 citations
Thyroid antibodies	127 citations
Febrile convulsions	27 citations
Familial data on congenital defects of the brain, spinal cord, feet, palate, and skull	19 citations
Immunological aspects of thrombocytopenia in man	21 citations

In September 1964 Dr. Charles L. Bernier, formerly of Chemical Abstracts and the National Institute of Child Health and Human Development, joined the Library and was assigned the responsibility of coordinating the evaluation of the MEDLARS system. All feedback reports are carefully evaluated, and suggestions for improvement forwarded for action. An analysis of the returns of services provided in 1964 suggests that MEDLARS is achieving a high level of relevance—an average of 74 percent of citations retrieved through these test searches has been called relevant by the requesters. We know of no other computerized system which has yielded such rewarding results at this stage of development.

The testing of relevance and recall rates in machine retrieval systems is one of the more active areas in information science. The Library is intensely interested in the application of formal testing methodologies to the MEDLARS system. However, the experimental and proposed methodologies need both development and further trial before application to such a large and complex system as MEDLARS.

In the meantime the Library has deferred the date when MEDLARS

will be open to all demand search requests until its manpower resources are sufficient to handle the load. The number of trained searchers on the Library staff now available is three. More are requested in the next fiscal year, and the Library has instituted a formal training program covering both indexing and search strategy. Experience has shown that searchers need to be thoroughly conversant with *MeSH* and its use, and experience in indexing has been considered essential.

While the Library is willing to accept a limited number of search requests submitted through other medical libraries for its continued purposes of testing and systems improvement, each service request should be negotiated in advance with the understanding that the requester supply critical analysis of the results. Hopefully in the fall of 1965 the Library may relax these requirements. In the meantime, NLM is continuing its efforts to strengthen *MeSH* for indexing, searching, and cataloging.

Indexing Improvement

The early testing of the system made apparent certain areas of weakness in the man-machine relationships involved in indexing with *MeSH*. One significant area was that of the adequacy of the subject heading system itself. *MeSH* had been derived within the Library principally from analyses of the published clinical literature and did not always reflect the terminological preferences of workers in specialized areas of basic medical science. Another weakness emerged in testing, notably the uneven distribution of the generic-specific character of the terminology in relation to the varying volumes of citations under specific subjects. Many of these weaknesses showed up only when searchers were called upon to formulate search strategies to answer complex inquiries put to the system by research scientists.

The area of *MeSH* development has been one of the most active in the entire Library. Early in the year, a meeting of advisers representing groups interested in the development of specialized controlled vocabularies and microthesauri in the medical sciences was called for the purpose of charting the interrelationship of a centralized terminological authority with the multiplicity of specialized vocabularies being developed for smaller systems purposes. Dr. Peter Olch, a trained pathologist and surgeon, joined the Library on July 1, 1964, to head a staff section devoted to the immediate strengthening of the Medical Subject Heading System and to its long-range development. Certain broad areas of science—dentistry; the behavioral sciences, including psychology and psychiatry; drugs and chemicals; and epidemiology—have been identified as priority areas for development, but the *MeSH* group of necessity had to turn its attention immediately to the preparation of the 1965

edition. This effort involved a critical review of over 600 provisional headings and included special attention, with the assistance of personnel from the American Dental Association, to the development of terminology covering the dental literature to facilitate the publication of a dental index described later.

The development of *MeSH* will be vigorously prosecuted. Advisers have been selected to assist the Library in developing *MeSH* terminology for the fields noted above. In addition, the Library has approached the National Research Council suggesting the establishment of a continuing advisory group on scientific terminology. In addition to the terminology itself, other aspects of *MeSH* constitute continuing problems on which the Library is working. The first relates to the development of tree structures, or formalized generic-specific patterns of terminology related to the categories. The efficiency and power of demand search procedures for MEDLARS depends heavily on the development and systems implementation of more of such structures.

In response to suggestions and criticism from users of *MeSH*, NLM has engaged in a study of "relationship indicators," or machine-acceptable means of subdividing broad subject concepts. For *Index Medicus* users these "indicators" may be near relatives of former subheadings.

Finally, the Library is working intensively on the application of *MeSH* to the mechanization of its book cataloging. In order to maintain consistency within the sexennial edition of the NLM *Catalog* to be published in 1966, book cataloging at the Library continues to be done on the basis of the 1964 *MeSH*. The Library maintains its commitment to employ the same subject headings for indexing as it does for book cataloging. At the same time, the implications of dedication to continuous revision of *MeSH* not only as it affects the Library's card catalog and those of other libraries, but as it places demands on the file-maintenance practices within the MEDLARS system itself, are being vigorously explored.

In addition to the strengthening of the subject heading system itself and its modification for specialized needs, evaluation of the searches performed suggested a degree of inconsistency in the use of headings by the indexers. Concurrent with the efforts to improve *MeSH*, therefore, the Library undertook to improve the practice of indexing. During the spring of 1964, a study and test of the Library's indexing conducted by Dr. William Spring and Mr. Boris Anzlowar suggested the adoption of a two-track system for indexing. The Library, with the advice of consultants, identified 792 journals as those most desirable to be indexed in increased depth, meaning an increase in the average number of headings assigned each paper. These were routed to senior indexers, while less

experienced indexers and trainees were assigned the less scientifically complex papers to be indexed in less depth. A significant accomplishment in the course of a year was the preparation of an indexing manual where for the first time many practices and decisions were codified. This manual will prove increasingly useful as the Library develops its indexer and searcher training program.

Recurring Bibliographies

The recurring bibliography, or periodic listing of citations pertinent to a given field of medical science selected from additions to the computer store, was one of the three basic objectives of MEDLARS. Recurring bibliographies are intended as selective dissemination of information to groups; composed by GRACE and published and distributed by national scientific societies or agencies of government, they constitute specialized indexes derived from a common MEDLARS base.

Progress in the establishment of recurring bibliographies has been slower than in any other area of MEDLARS. The reasons are threefold: first, insufficient personnel in the Library and cooperating groups to give full time to the detailed planning and development; second, difficulties experienced by representatives of the interested groups in defining fields of interest in terms of *MeSH;* and, third, the basic problem of issuing a subject bibliography without postediting.

The last problem can be summarized by stating that all citations destined for a recurring bibliography are *pretagged* at the time of entry into MEDLARS. There is no possibility of editing before their composition by GRACE. A percentage of nonrelevance is inevitable, yet the relevance requirements for published bibliographies are considerably higher than for demand searches, where editing is possible. The Library is presently considering the alternative of producing these bibliographies through recurrent demand searches.

Despite these difficulties, the Library in cooperation with several groups made progress during the year, and issues of several recurring bibliographies have been or are about to be published. These are:

Rheumatology Bibliography. Under a grant from the National Institute of Arthritis and Metabolic Diseases, the American Rheumatism Association employed a full-time investigator working with NLM to develop the parameters of this recurring bibliography. Noteworthy was the development of a microthesaurus for the field, showing relationships of terminology to *MeSH.* Three trial issues of an Index to Rheumatology, intended as a biweekly publication of the American Rheumatism Association, have been prepared for review and criticism.

Cerebrovascular Bibliography. A joint committee representing the

National Institute of Neurological Diseases and Blindness and the National Heart Institute, working with the cooperation of NLM, has produced several trial issues of this bibliography for critical review. Viewed as a pilot study, this bibliography has yet to be issued by the recurring bibliography mechanism.

Fibrinolysis and Blood Coagulation Bibliography. Under the auspices of a Committee on Thrombolytic Agents of the National Heart Institute, this bibliography has developed to a point where trial runs may be expected early in 1965.

Index to Dental Literature. Described elsewhere, this index has had strong support from the American Dental Association. Intended as a quarterly, cumulating through the year to an annual, a first issue is planned for April 1965.

In addition, recurring bibliographies are in process or under negotiation in the following fields: digestive diseases, electromyography and electrodiagnosis, sudden death in infants, medical education, and smoking and health.

The Bibliography of Medical Reviews, technically a recurring bibliography, since it consists of pretagged citations, presented no major complications, and the 1964 issue was composed by GRACE for publication by the Library. In addition, the *Abridged Index Medicus,* mentioned elsewhere, may be considered a recurring bibliography.

Cooperative Agreements

As the Library developed its own plans for services and products from MEDLARS, other agencies, public and private, became interested in specialized services. The result was a series of agreements committing NLM to the provision of specialized services, on the one hand, and the participating agency, on the other hand, to the contribution of manpower or funds, so badly needed by the Library to expand its indexing and searching capability.

An agreement was signed with the Veterans Administration in July 1964 through which the VA agreed to supply a systems analyst and two indexer-searchers. The VA wishes to study the system in anticipation of its possible adoption and establishment in that agency. An agreement was signed with the Food and Drug Administration in February 1964 for the provision of MEDLARS services; searches on drugs have been performed regularly, and the Library anticipates strengthening the system and increasing its services in this important field. Similarly the Library has been exploring a cooperative agreement with the Agency for International Development, whereby MEDLARS search services would be made available to the overseas health missions of that agency.

On the private front, the Library signed an agreement with the American Dental Association in August 1964 for the purpose of preparing an index to the dental periodical literature as a special product of MEDLARS. With two dentists employed by the ADA working with the NLM staff, NLM indexing was expanded to cover 100 core dental journals. New subject headings were developed to provide more specificity to the indexing of the dental literature. To this indexing will be added that part of the total MEDLARS indexing pertaining to dentistry and the additions supplied by ADA staff in Chicago, and the resulting index will be composed by GRACE and published by the ADA.

The possibility of publishing an index to nursing literature was also explored by the Library and the American Journal of Nursing Company in the course of the year.

The agreement with the American Medical Association for the *Cumulated Index Medicus* was continued during the year. The film copy of the *CIM* for 1964, produced by MEDLARS and composed by GRACE, involved approximately 6,000 pages of composition; it was completed and sent to the AMA on December 15.

MEDLARS Decentralization

The need for decentralizing MEDLARS was foreseen in the initial contract with the General Electric Company, which specified compatibility with certain secondary objectives, including "requirements for a national, decentralized, medical bibliographic system." Behind this concept lay two considerations: first, that demands for machine search would outstrip machine and human resources available to satisfy them from a central service facility, and, second, that exploitation of a centrally produced set of tapes was preferable economically to the creation of multiple incompatible tape systems. An analogy may be made to the publication and use of a centrally prepared index: decentralization of MEDLARS tapes equates to the publication of the *Index Medicus* in magnetic tape format.

Technical difficulties were twofold. First, the compressed citation file (CCF) containing the references to be distributed was in a physical format acceptable only by a Honeywell computer; second, the directions to be given the computer for processing searches (programs) have been expressed in a logical language (ARGUS) which only the Honeywell computer could use. Tape conversion can be accomplished by machine, but the rewriting of computer programs must be done by humans.

Accordingly, as a first step toward decentralization, the Library contracted with UCLA to reprogram MEDLARS from ARGUS to COBOL, a language accepted by IBM computers of the 7090 series. The contract

was signed in November, and the reprogramming should be completed by July 1965. The Library has acquired a Honeywell 200, which will have the ability to copy and convert tapes. This effort, which will culminate in an ability at UCLA to use MEDLARS tapes for searching, marks the first stage of a planned program for MEDLARS decentralization to other universities and research centers.

The COBOL programs developed at UCLA and the tapes converted at NLM will be available for the planned second stage, to be implemented after July 1965. A small number of university centers will be selected from the forty which have expressed an interest to date. Support will be in the form of a contract, with NLM meeting the costs incurred in providing machine searches to a mutually agreed public. This phase represents a period of training, demonstration, and field testing, and experience gained will be fed back to NLM to strengthen the system.

At its meeting in November 1964, the NLM Board of Regents considered and approved policies underlying decentralization, including criteria to be followed in the selection of institutions for support in this phase. Among the criteria are the character of the existing computer resources, the library resources behind them, the size and character of the populations to be serviced, and geographic distribution of the search centers.

Some confusion has arisen over the relationship between the MEDLARS search centers to be supported in this phase and "regional libraries" as discussed by members of the Medical Library Association. The selection of a library as a MEDLARS search center is not predicated on a willingness to provide regional library service as a unit of a national network. There are many unknown factors in the design of a national medical library network, in which certain institutions might undertake a regional function, and NLM hopes to support studies in this area through its Extramural Program. It is difficult to conceive of regional libraries, if such are to be established, without MEDLARS search capability; on the other hand, local MEDLARS search centers, servicing various populations, can be established without prejudice to the design of a system of regional medical libraries.

Assuming the passage of the appropriation legislation, selection of institutions for this second phase will be made for confirmation of the NLM Board of Regents at its June 1965 meeting, and contracts will be negotiated after July 1, 1965.

Support for further decentralized search centers is planned in subsequent fiscal years, again related to the needs for training, demonstration, and test. Concurrently, NLM plans to make tapes available at cost to other users.

MEDLARS Internationalization

Just as the published *Index Medicus* has an international use, so it may be anticipated that the Compressed Citation File of MEDLARS, which in effect constitutes the *Index Medicus* in magnetic tape form (with the added power of increased depth of indexing and more search parameters), will have international interest.

Following the MEDLARS session at the Second International Congress on Medical Librarianship, a number of countries indicated their interest in acquiring MEDLARS tapes. While it would be premature to list these overtures, it can be reported that the Swedish Medical Research Council has sponsored the training of Dr. Kjell A. H. W. Samuelson, and the new Ministry of Education and Science in the United Kingdom is also sending a representative for orientation early in 1965.

Other countries as well as international organizations have sent representatives to NLM to review the capability of MEDLARS, and it may be anticipated that as MEDLARS becomes established in other institutions in the United States there will be increasing interest in other countries for the establishment of MEDLARS search centers overseas.

MEDLARS in Technical Processes

It is perhaps ironical that the area of computer applications most actively developed by other libraries—that of procurement, serial record maintenance, and cataloging—was designated a secondary objective for MEDLARS. Important as these functions are to NLM and other research libraries, bibliographical publication and literature search were the primary objectives.

Late in 1963 a decision was made to accomplish these secondary objectives "inhouse," rather than under contract, and Mr. Irvin Weiss joined the staff of the Data Processing Division for the express purpose of conducting systems analysis, on February 2, 1964. Concurrently, NLM established liaison with others working in the field, and a series of meetings was held with the pioneering Harvard-Yale-Columbia group.

NLM approached the problem with an open mind, determining only to accomplish a thorough systems analysis and flow charting, without commitment in advance to mechanization. Four areas—searching, procurement, serial record, and cataloging—were considered as interrelated modules of a system, and flow charting proceeded.

The systems analysis and flow charting were completed in December 1964, and the decision was made to mechanize and to assign priority to the cataloging function. A primary reason was the incompleteness of MEDLARS without representation of the monographic literature and of

analytics from composite works. Lending urgency was the need for decision on the means of preparation of the annual NLM book *Catalog* to succeed the publication by the Library of Congress of the manually prepared sexennial (not quinquennial) cumulation in the spring of 1966. Programming on an interim cataloging module started in January 1965.

The Library's objectives in computerizing its cataloging are: (1) the production of a biweekly bulletin for distribution to other libraries to assist them in their acquisitions and cataloging; (2) the publication of an annual catalog, with a five-year cumulation; (3) the production of sets of cards for filing in its card catalog; and (4) the input of catalog entries for appropriate books and for analytics from composite works into MEDLARS.

The needs of other libraries must be known as the Library defines its objectives, and, to represent these needs and to provide expert advice, the Library has established several consultancies.

In the area of subject headings, a number of questions are to be resolved. By original design, the Library is committed to the use of a single set of headings for both the indexing of periodical articles and the cataloging of books. The 1963 *MeSH* was intended for use in MEDLARS, where the associational powers of the computer are used to gain specificity of meaning. Missing the display of a subheading structure (although the use of subheadings is still possible for cataloging), some librarians protested. The Library sponsored an informal discussion of the problem at the San Francisco meeting of the MLA and agreed to produce a *MeSH* User Group Bulletin, the first number of which will be published early in 1965. The Bulletin is intended as a forum for the discussion of new medical subject headings and desirable changes.

In addition, the Library is investigating changes which must be made to give the computer system increased capability of using subheadings and/or relationship indicators.

MEDLARS Mark III

While much energy has been expended in making MEDLARS operational and in resolving the day-to-day systems and service problems, research opportunities leading to the development of the next generation of a MEDLARS system have been sought and exploited. Basic work done outside the Library is followed and its possible application assessed; developments in other systems are similarly studied. Among the areas receiving continuing study are:

Input. Rather than input typists converting information manually

supplied by indexers to machine form, can there be a form of computer-aided direct input by the indexers themselves?

Natural language indexing. Can the computer efficiently perform editing on uncontrolled terminology abstracted by the indexer and print out for study by lexicographers terms not in the system?

"Project Specify." A committee chaired by Dr. Charles Bernier is reviewing the long-range problem of increasing specificity of indexing terms as the volume of citations in MEDLARS increases.

Graphic image storage and retrieval. A group from the National Bureau of Standards is working in the Library with the immediate purpose of streamlining the Library's photoduplication services and the long-range objective of testing the feasibility of a graphic image storage system which might be associated with MEDLARS.

Conclusion

MEDLARS is one of a very few large operational computer-based information retrieval systems in existence. Its first year has been marked by challenge, trial, and further development. Progress toward realizing its full service potential has been slow because of NLM manpower limitations, but its performance as a system leaves no question as to its fundamental soundness.

As a system also, it has proven that it is not and cannot be static. The continuing quest for perfection brings change and results in system growth.

As MEDLARS moves into its second year, it is a stronger, better system than it was a year ago and is prepared to take on the wider challenge of open service and decentralization. Feedback and constructive criticism are invited. Such stimuli will assist NLM in its effort to improve its MEDLARS services.

REFERENCE

1. Adams, Scott. MEDLARS and the library community. Bulletin 52: 174, Jan. 1964.